PROTOCOL TO EVALUATE THE STATE OF THE SCIENCE FOR TRANSGENERATIONAL INHERITANCE OF HEALTH EFFECTS

Project Leader: Vickie R. Walker, Office of Health Assessment and Translation (OHAT), DNTP

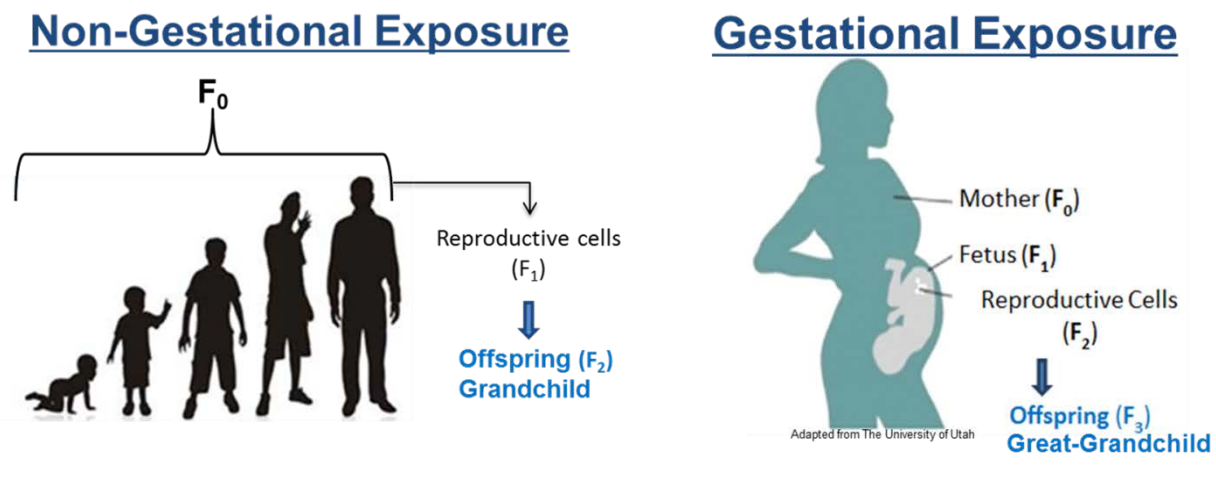
Summary: OHAT is conducting a state-of-the-science evaluation to examine the extent of the evidence for transgenerational inheritance of health effects associated with exposure to a wide range of stressors (e.g., environmental chemicals, drugs of abuse, nutrition and diet, pharmaceuticals, infectious agents, or stress) in humans or non-human animals. The state-of-the-science evaluation will provide a critical analysis of the literature related to transgenerational inheritance of health effects with a goal of identifying areas of consistency and areas of uncertainty as well as data gaps and research needs.

BACKGROUND AND SIGNIFICANCE

Background

There is evidence that early life exposures can lead to disease outcomes much later in life (Barker *et al.* 2002). The traditional dogma suggests that negative effects of these early life exposures do not carry over to subsequent generations, and thus these generations are unaffected by the exposure history of their parents and grandparents. However, there are reports of exposures, which, in certain cases, appear to have far-reaching consequences that affect multiple generations beyond the original exposure generation (reviewed in Grossniklaus *et al.* 2013, Aiken and Ozanne 2014). This phenomenon of health effects in offspring that were not themselves exposed to a chemical or non-chemical stressor is known as "transgenerational inheritance." **Figure 1** outlines the transgenerational inheritance exposure paradigm. Non-gestational exposure occurs in the F_0 generation but then stops and is not continuous across generations. However, the germ cells (and thus the F_1 generation) are exposed. Gestational exposure occurs during pregnancy and therefore exposures are to the pregnant female (F_0), the fetus (F_1), and the germ cells developing with the fetus (F_2); the F_3 generation is the first generation

Figure 1. Transgenerational inheritance exposure paradigm



not exposed directly. Multigenerational studies with continuous exposure to the stressor across all generations (F_0, F_1, F_2, etc.) would not be considered a transgenerational study design. To evaluate true transgenerational effects in these two models, the health effect is evaluated at a minimum in the F_3 or later generation for gestational exposure and the F_2 or later generation for non-gestational exposure, the first unexposed generations for each model.

Significance

The state-of-the-science evaluation will pull together a literature base that is challenging to identify and serves as a background document for the field of research considering transgenerational health effects. Data management will be conducted in a manner that permits sharing of data extraction files with the public and other agencies. The evaluation will also illustrate how internal validity or risk-of-bias assessment could be applied to studies with a transgenerational study design using subsets of identified human and animal[1] studies as examples. The risk-of-bias assessment can serve to demonstrate potential sources of bias in study design, conduct, and reporting that could be considered by researchers in ongoing and future studies in the field of transgenerational inheritance.

OVERALL OBJECTIVE AND SPECIFIC AIMS

The overall objective of this evaluation is to develop a state-of-the-science evaluation to examine the extent of the evidence for transgenerational inheritance of health effects associated with exposure to a wide range of stressors (e.g., environmental chemicals, drugs of abuse, nutrition and diet, pharmaceuticals, infectious agents, or stress) in humans or animals. This state-of-the-science evaluation will provide a critical evaluation of the literature related to transgenerational inheritance of health effects with a goal of identifying areas of consistency and areas of uncertainty as well as data gaps and research needs.

Specific Aims

1. Identify literature within PubMed[2] in humans and animals utilizing a transgenerational study design that assess a wide range of stressors (e.g., environmental chemicals, drugs of abuse, nutrition and diet, pharmaceuticals, infectious agents, stress) with no restriction on the type of health outcome.
2. Extract data on potential health effects from relevant studies (data extraction files will be shared upon release of final report).
3. Provide an inventory of included studies, focusing on type of exposure and health outcome for both human and animal studies.
4. Prepare a brief summary of the study results, grouped by exposures for which the same health outcome is reported, to identify areas of consistency and areas of uncertainty.
5. Assess risk of bias of individual human and animal studies from a subset of the relevant references to illustrate how risk-of-bias assessment could be applied to studies of

[1] The terms “animal” or “animals” in this protocol refer to non-human animals.

[2] Note: a search of PubMed identified over 50,000 references that have been searched and screened for relevance and eligibility for this review. NTP considered broadening the database coverage with a similar search of TOXLINE and SCOPUS. However, due to the low specificity of the search (there are no major subject heading search terms for “transgenerational”), inclusion of those databases would expand the number of references 2- or 3-fold to between 100,000 and 200,000 references. Thus, we limited our search to PubMed.

transgenerational study design. The subsets will be studies from the 2 health effects with the most extensive bodies of evidence for both human and animal studies.

6. Outline key issues or data gaps that could be addressed in future research of transgenerational inheritance of health effects.

PECO Statement

To address our overall objective we developed a PECO (**P**opulation, **E**xposure, **C**omparators and **O**utcomes) statement (**Table 1**), which is used as an aid to develop the specific research questions, search terms, and inclusion/exclusion criteria for our systematic review (Higgins *et al.* 2011).

Table 1. Human and animal PECO (Population, Exposure, Comparators and Outcomes) statement

PECO Element	Evidence
Population	Human or animal (whole organism) without restriction based on age, sex, or lifestage at exposure or outcome assessment
Exposure	Any exposure/stressors at any life stage as long as the study design (i.e. outcome measurement stage) is was transgenerational in nature (**Figure 1**).
Comparators	*Humans:* A comparison population exposed to lower levels (or no exposure/exposure below detection levels) of the stressor than the more highly exposed subjects *Animals*: Comparable animal populations exposed to vehicle-only treatment in experimental studies or a comparison animal population exposed to lower levels (or no exposure/exposure below detection levels) of the stressor than more highly exposed animals in wildlife/farm animals *Note: the comparison groups are defined at the time of the exposure and therefore apply to the F0 generation (which in a gestational exposure would include exposure of the offspring (F1) and its gametes (F2) [see figure 1].*
Outcomes	Inherited diseases were excluded. No other restrictions on health outcome or effect measures

The overall objective, PECO statement, and strategy to synthesize study results were based on a series of problem formulation steps that included: (1) deliberation with NTP staff and consultation with scientists at other Federal agencies; (2) consultation with an evaluation design team with expertise in reproductive and developmental toxicology, epidemiology, epigenetics, systematic review, and information science; (3) a public request for information in the Federal Register [88 FR 26646 (May 7, 2013)], (4) consideration of the extent of information available for specific types of exposures, and (5) future consideration of any comments received on the evaluation protocol following anticipated posting in May, 2015. More details about problem formulation can be found below.

Key Questions

The overall objective of this state-of-the-science evaluation is to investigate the evidence for transgenerational inheritance of health effects. This objective will be answered by addressing the following key questions.

Table 2: Key Questions (KQ)	
KQ1	What is the extent of the evidence of human and animal studies for exposures with a transgenerational study design?, This includes:, a) Identification of exposures (e.g., environmental chemicals, drugs of abuse, nutrition and diet, pharmaceuticals, infectious agents, or stress) and outcomes (e.g. reproductive, developmental, neurological and behavioral) that have been studied for transgenerational inheritance. b) Organize and group the literature to identify bodies of evidence for which there are multiple studies of a given exposure and effect (i.e., the same health outcome was investigated for the same exposure).
KQ2	What are the areas of consistency and areas of uncertainty in the identified bodies of transgenerational inheritance literature and whether this information suggests data gaps or research needs?

METHODS

Step 1. Problem Formulation

Rationale

This evaluation of transgenerational inheritance of health effects addresses a long-standing topic of interest for NIEHS (e.g., NIEHS 2012) and also overlaps with the institute's interest in epigenetics, which emerged as a key theme in the NIEHS strategic plan (http://www.niehs.nih.gov/about/strategicplan/). OHAT received a nomination to evaluate the evidence for epigenetic effects and is currently developing an approach to evaluate the epigenetic literature. As a complement to the epigenetics project, the evaluation design team decided that a state-of-the-science report summarizing the evidence for the general phenomenon of transgenerational inheritance would be helpful because the evaluation could focus on the evidence for potential health effects rather than the mechanisms. This survey and collection of the health effects evidence for transgenerational inheritance will identify the exposures and health outcomes that have been studied to date. Further, it will identify larger bodies of evidence with multiple reports investigating potential effects of a given exposure on the same health endpoint. Thus a state-of-the-science document and discussion of areas of consistency and areas of uncertainty in the identified bodies of evidence will be an important first step to evaluating the phenomenon of transgenerational inheritance. The report may assist ongoing investigations of epigenetics literature because many of the nutritional and chemical stressors implicated in potential transgenerational inheritance of health effects have the capacity to alter the epigenome (e.g., Guerrero-Bosagna and Skinner 2012). Identifying the intersection of substances addressed in both the epigenetics and transgenerational inheritance literature will help with the interpretation of the current literature, as well as the development of research strategies moving forward.

Problem Formulation Activities

In January 2013, OHAT solicited input from scientists at the NIEHS and other federal agencies on the NTP Executive Committee[3] on the proposed evaluation. The committee supported OHAT moving forward with the project and encouraged OHAT to keep the range of stressors broad, including drugs of abuse and infectious agents. In May 2013, OHAT published a request for information (RFI) in the Federal Register [88 FR 26646 (May 7, 2013)]. The RFI requested public input on a draft search strategy that could be used to conduct an initial inventory of the literature, potential areas of focus for the evaluation, identification of unpublished, ongoing or planned studies related to transgenerational inheritance, as well as identification of scientists with expertise or knowledge in this area who might be consulted during the course of OHAT's evaluation. Similar to the Executive Committee comments, the 5 public comments received in response to the RFI suggested that we keep the exposures and health outcomes broad; although, there was a suggestion that we emphasize the importance of potential direct germline effects. The commenter suggested categories of potential stressors that we have used for the evaluation (e.g., pharmaceuticals, radiation, drugs of abuse, environmental chemicals); and we have supplemented these with the addition of categories for infectious agents, diet and stress. In addition, several search terms were suggested by commenters that apply to potential mechanisms (e.g., methylation, histone, epigenetic) or life stages (prenatal, *in utero*, gestational). As stated in the rationale, this state-of-the science evaluation is focused on collecting the evidence for transgenerational inheritance of health effects independent of mechanism and therefore the search terms for potential mechanisms were not included. Similarly, no modifications of the search strategy were made to reflect the life stage-specific suggestions, as this would narrow the literature retrieved. As stated in the PECO statement, the populations of interest are humans or animals without restriction based on age, sex, or lifestage at exposure or outcome assessment. Several commenters identified scientists with expertise relevant to this topic.

Consideration of key scientific issues

Several key scientific issues were identified during the initial consideration of the literature. A summary of those issues and how OHAT will address them in the evaluation are summarized below.

1. **The challenges of consistency in reporting studies of "transgenerational" inheritance.** Clear and uniform definitions of the term "transgenerational" have not been defined consistently in the literature and presents challenges to accurately identify the various generations and their exposures. In addition to this lack of consistency in terminology used to report transgenerational study designs, there is also a lack of a specific Medical Subject Headings (MeSH) term to identify and index these study types. OHAT conducted a broad literature search of PubMed that was biased towards not missing studies utilizing a transgenerational design, even though as a result many irrelevant studies were retrieved.

2. **Consideration of animal studies of known mutagenic chemicals.**

[3] Consumer Product Safety Commission (CPSC), Department of Defense (DoD), Environmental Protection Agency (EPA), Food and Drug Administration (FDA), National Cancer Institute (NCI), National Center for Environmental Health/Agency for Toxic Substances and Disease Registry (NCEH/ATSDR), National Institute of Environmental Health Sciences (NIEHS), National Institute for Occupational Safety and Health (NIOSH), Occupational Safety and Health Administration (OSHA) http://ntp.niehs.nih.gov/go/163

Studies of known mutagens will not be excluded from the evaluation because the goal is to survey all exposures that fit the transgenerational study design. For most exposures, the mechanism of action is unclear. However, for known chemical mutagens, mutagenicity would be one of the possible mechanisms for transgenerational inheritance. Although mechanisms will not be a focus for of this evaluation, the data on exposures with a known mutagenic mechanism of action will be discussed separately in the report and the analysis of mutagens will include a discussion of the mechanism.

3. **Consideration of human studies of migration and socioeconomic status stressors.** Tracking potential effects over multiple human generations is particularly challenging due to potential confounding over multiple generations. This is especially true for studies of socioeconomic impact and migration, which may occur in each generation. For this reason, we will focus our analysis on exposures or stressors that were more likely to occur in a discrete period of time and were not continuous (e.g., radiation).

Step 2. Search For and Select Studies for Inclusion

Literature Search Strategy

Search terms were developed by an informationist familiar with systematic review methodologyto identify all relevant published evidence indexed in the PubMed database (MEDLINE) that addresses the key questions on transgenerational inheritance of health effects potentially associated with a wide range of exposures (e.g., environmental chemicals, drugs of abuse, nutrition and diet, pharmaceuticals, infectious agents, or stress). Because there are no (MeSH) for transgenerational effects or transgenerational inheritance, our search strategy was developed using terminology that describes a transgenerational study and therefore would potentially be used to report the topic of transgenerational inheritance in the literature. The topic of transgenerational inheritance has been reported in various ways in the published literature and required the development of a search strategy that addresses key terms associated with the various concepts of transgenerational inheritance. Our strategy included key words to identify the concepts of transgenerational, multigenerational or intergenerational studies as well as terms to address successive generations (search terms presented Appendix 1).

A test set of relevant studies was used to ensure the search terms retrieve 100% of the test set. No language restrictions or publication year limits will be imposed in the search terms (see Table 3 for inclusion/exclusion search terms) and the literature search will be updated for a final time approximately 90-120 days prior to peer-review and the new studies identified will be evaluated and incorporated into the body of literature for this evaluation.

Applying this search strategy to PubMed identified over 50,000 references. NTP considered broadening the database coverage with a similar search of Web of Science and SCOPUS. However, due to the low specificity of the search, inclusion of those databases would expand the number of references 2- or 3-fold to between 100,000 and 200,000 references (e.g., adding approximately 79,500 references from Web of Science and 129, 500 references from Scopus). Thus, we limited our search to PubMed for this state-of-the-science report.

Searching other resources

We will use the following methods to find studies that would not be identified through the electronic searches. Studies will be evaluated using the same inclusion and exclusion criteria as used for screening records retrieved from the electronic search. Relevant studies identified through these steps will be marked as "provided from other sources" in the study selection flow diagram.

- Hand searching the reference lists of all included studies after the full text review.
- Hand searching the reference lists of relevant reviews, commentaries, or other non-research articles identified during the initial search. Commentaries or letters on specific studies are also reviewed to see if they contain content that should be noted during data extraction.
- Studies identified by the public when the initial list of included studies is posted on the OHAT website and distributed via NTP and NIEHS DERT list serves to solicit input on the list of included studies (approximately 60-90 days prior to peer-review; studies identified within 30 days of posting will be considered for inclusion) or during the public comment period when the draft state-of-the-science report is released for public comment (approximately 45-60 days prior to peer-review).

Unpublished data

Although the search of the PubMed database will only retrieve published literature, unpublished data may be identified and submitted by the public. NTP only includes publicly accessible, peer-reviewed information in its evaluations. If a study is identified that may be critical to the evaluation and is not peer reviewed, the NTP's practice is to obtain external peer review if the owners of the data are willing to have the study details and results made publicly accessible. The peer review would include an evaluation of the study similar to that for peer review of a journal publication. The NTP would identify and select two to three scientists knowledgeable in scientific disciplines relevant to the topic as potential peer reviewers. Persons invited to serve as peer reviewers would be screened for conflict of interest (COI) prior to confirming their service. In most instances, the peer review would be conducted by letter review. The study authors would be informed of the outcome of the peer review and given an opportunity to clarify issues or provide missing details. OHAT would consider the peer review comments regarding the scientific and technical evaluation of the unpublished study in determining whether to include the study in its evaluation. The study and its related information, if used in the OHAT evaluation, would be included in the systematic review and publicly available. OHAT would acknowledge via a note for the report that the document underwent external peer review managed by the NTP, and the names of the peer reviewers would be identified. Unpublished data from personal author communication can supplement a peer-reviewed study, as long as the information is made publicly available.

Screening Process

References retrieved from the literature search will be screened for relevance and eligibility using DistillerSR®, a web-based, systematic-review software program with structured forms and procedures to ensure standardization of the process[4]. Search results will first be consolidated in Endnote reference management software and duplicate articles will be removed prior to uploading the references into DistillerSR®.

[4]DistillerSR® (http://systematic-review.net/) is a proprietary project management tool for tracking studies through the screening process and storing data extracted from these studies using user-customized forms.

Evidence Selection Criteria

In order to be eligible for inclusion, studies must comply with the type of evidence specified by the PECO statement (**Table 1**). Inclusion and exclusion criteria based on the PECO statement are detailed in **Table 3**; these criteria are used to screen articles for relevance and eligibility at both the title-and-abstract and full-text screening stages. In addition to criteria defining the relevant population, exposure, comparator, and outcomes, **Table 3** defines criteria for relevant publications types (e.g., the report must contain original data). Studies that do not contain original data (review articles or reports that do not include health effects data) will be categorized as supportive material that may contain relevant background information (e.g., reviews or commentaries that discuss transgenerational study design or studies that report relevant exposure or metabolism data) that could be useful when evaluating areas of consistency and uncertainty in the bodies of evidence from the included studies.

Multiple publications of same data

Multiple publications with overlapping data for the same study (e.g., publications reporting subgroups, additional outcomes or exposures outside the scope of an evaluation, or longer follow-up) are identified by examining author affiliations, study designs, cohort name, enrollment criteria, and enrollment dates.

Table 3. Inclusion and Exclusion Criteria to Determine Study Eligibility		
	Inclusion Criteria	**Exclusion Criteria (or blank if none)**
Population (Human Studies or Experimental Model Systems)		
human	• No restrictions on sex, age, or lifestage at exposure or outcome assessment	
non-human animal	• No restrictions on sex, age, species, or lifestage at exposure or outcome assessment	• Studies in non-animal organisms (e.g., plants, fungi)
Exposure		
human or non-human animal	• Exposure required, but no restrictions on the type of exposure	• *In vitro* exposure studies
Comparators		
human	• Humans exposed to lower levels (or no exposure/exposure below detection levels) of the stressor than more highly exposed subjects	
non-human animal	• ***For experimental studies***: study must include vehicle or untreated control group • ***For wildlife or observational studies***: animals exposed to lower levels (or no exposure/exposure below detection levels) of the stressor than the more highly exposed subjects	
Outcomes		
Human or non-human animal	• Health outcome required in the relevant generation depending on the timing of exposure (see Figure 1), but no restrictions on the type of health outcome reported	
Publication Type (e.g., specify any language restrictions, use of conference abstracts, etc.)		
Human or Non-human animal	• Report must contain original data • Report must be in English language	• Articles with no original data (e.g., editorial or review*) • non-English** • Studies published in abstract form only (grant awards, conference abstracts) • Retracted articles

*Relevant reviews are used as background and for reference scanning.

**Because this is a state-of-the-science document the non-English language studies will be excluded to minimize the additional time and cost associated with reviewing these references.

If necessary, study authors will be contacted to clarify any uncertainty about the independence of two or more articles. OHAT will include all publications on the study, select one study to use as the primary, and consider the others as secondary publications with annotation as being related to the primary record during data extraction. The primary study will generally be the publication with the longest follow-up, or for studies with equivalent follow-up periods, the study with the largest number of cases or the most recent publication date. OHAT will include relevant data from all publications of the study, although if the same outcome is reported in more than one report, OHAT will include a single instance of the data (and avoid more than one, i.e., duplicate instances of the data).

Title/Abstract Review

Screeners will be trained using project-specific written instructions that reflect the criteria outlined in **Table 3** with an initial pilot phase undertaken to improve clarity of the inclusion and exclusion instructions and to improve accuracy and consistency among screeners. If changes to the inclusion criteria are made based on the pilot phase, they will be documented in a protocol amendment along with the date modifications were made and the logic for the changes. Trained screeners from the evaluation design team will then conduct a title and abstract screen of the search results to determine whether a reference meets the inclusion or exclusion criteria. All references will be independently screened by two screeners (one of which will be the project lead, who will screen all references). Studies that are not excluded based on the title and abstract will be screened through a full-text review. In case of screening conflicts, screeners will independently review their screening results to confirm the inclusion/exclusion decision and, if needed, discuss discrepancies with the other screeners. If a true disagreement exists between screeners, the study passes to the full-text review.

Full-Text Review

After completion of the title/abstract screen, full-text articles will be retrieved[5] for those studies that either clearly meet the inclusion criteria or where eligibility to meet the inclusion criteria is unclear. Full-text review will be independently conducted by two screeners that participated in the title/abstract screening (again, one of which will be the project lead, who will screen all references). True disagreements will be resolved by discussion with the project lead. If disagreements are not immediately resolved, consultation with other members of the evaluation design team and technical advisors will be used to reach consensus.

Tracking study eligibility and reporting the flow of information

The main reason for exclusion at the full-text-review stage will be annotated and reported in a study selection flow diagram in the final report (using reporting practices outlined in Moher *et al.* 2009). The following reasons for exclusion will be documented: (1) is a review, commentary, or editorial with no original data; (2) is not a transgenerational study design; (3) lacks exposure information; (4) lacks health outcome information; (5) only data on non-animal organisms (e.g., plants); (5) Non-English article or (5) is a conference abstract, grant application/registration, or thesis/dissertation.

Release of the list of included and excluded studies

The list of included and excluded studies will be posted on the OHAT website (http://ntp.niehs.nih.gov/go/evals) once screening has been completed and prior to completion of the draft OHAT state-of-the-science report.

[5]OHAT will initially attempt to retrieve a full-text copy of the study using an automated program, such as QUOSA, when possible, and NIH library services (NIH subscriptions and interlibrary loans). For publications not available through NIH, OHAT will search the Internet and/or may attempt to contact the corresponding author. Studies not retrieved through these mechanisms are excluded and notated as “not available.”

Step 3. Data Extraction

Data Extraction Process and Data Warehousing

Data extraction will be managed with structured forms and stored in a database format using Health Assessment Workspace Collaborative (HAWC) [6], an open source and freely available web-based interface application. Data extraction elements are listed separately for human and animal studies in **Appendix 2-3**. The data extraction results for included studies will be visualized and made publicly available in Excel format upon publication of the final state-of-the-science report.

The extracted data will be used to help summarize study designs and findings, facilitate assessment of risk of bias. The number of elements or collection of information on a specific element may be revised following the identification of important study details from individual studies included in the review. Data extraction will be performed by one member of the evaluation team (or contract support) and checked by a second member for completeness and accuracy. Any discrepancies in data extraction will be resolved by discussion or consultation with a third member of the evaluation team if the discrepancy is not immediately resolved. Information that is inferred, converted, or estimated during data extraction will be annotated, e.g., using brackets [n=10]. OHAT will attempt to contact authors of included studies to obtain missing data considered important for evaluating key study findings (e.g., level of data required to conduct a meta-analysis; note this is a standard to ensure clear reporting and this project will not include a meta-analysis). The evaluation report will note that an attempt to contact study authors was unsuccessful if study researchers do not respond to an email or phone request within one month of the attempt to contact.

Step 4. Quality Assessment of Individual Studies

Internal validity or risk of bias will be assessed for individual studies using a tool developed by OHAT that outlines a parallel approach for evaluating risk of bias from human or animal studies to facilitate consideration of risk of bias across evidence streams with common terms and categories. The risk-of-bias tool is comprised of a common set of 11 questions that are answered based on the specific details of individual studies to develop risk-of-bias ratings (using the four options in **Table 4**) for each question. Study design determines the subset of questions that should be used to assess risk of bias for an individual study (**Table 5**). For example, the subset of risk-of-bias questions applicable to all of the experimental study designs includes a question on randomization of exposure that would not be applicable to observational study designs. Therefore, a similar set of questions are used across experimental study designs (experimental animal and human controlled trials).

Studies are independently assessed by two assessors who answer all applicable risk-of-bias questions with one of four options in **Table 4** (answers from CLARITY Group at McMaster University 2013) following pre-specified criteria detailed in **Appendix 4** for the most common exposures encountered during the initial literature search: radiation and holocaust experience (families followed over generations subsequent to an individual experiencing the holocaust in an earlier generation) for the human studies and vinclozolin and radiation for the animal studies. The criteria describe aspects of study design, conduct, and reporting required to reach risk-of-bias ratings for each question and specify

[6] HAWC (Health Assessment Workspace Collaborative): A Modular Web-based Interface to Facilitate Development of Human Health Assessments of Chemicals (https://hawcproject.org/portal/).

factors that can distinguish among ratings (e.g., what separates “definitely low” from “probably low” risk of bias).

Table 4: Answers to the Risk-of-Bias Questions Result in One of Four Risk-of-Bias Ratings	
++	***Definitely Low*** *risk of bias*: There is direct evidence of low risk-of-bias practices
+	***Probably Low*** *risk of bias*: There is indirect evidence of low risk-of-bias practices **OR** it is deemed that deviations from low risk-of-bias practices for these criteria during the study would not appreciably bias results, including consideration of direction and magnitude of bias
– NR	***Probably High*** *risk of bias*: There is indirect evidence of high risk-of-bias practices (indicated with “-“) **OR** there is insufficient information provided about relevant risk of bias practices (indicated with “NR” for not reported). Both symbols indicate probably high risk of bias.
––	***Definitely High*** *risk of bias*: There is direct evidence of high risk-of-bias practices

The instructions and detailed criteria are tailored to the specific evidence stream and type of human study designs. Risk of bias will be assessed at the outcome level because study design or method specifics may increase the risk of bias for some outcomes and not others within the same study.

Risk-of-Bias Assessment Process

Assessors will be trained using the criteria in **Appendix 4** with an initial pilot phase undertaken to improve clarity of criteria that distinguish between adjacent ratings and to improve consistency among assessors. All team members involved in the risk-of-bias assessment will be trained on the same set of studies (i.e. 3-5 studies) and asked to identify potential ambiguities in the criteria used to assign ratings for each question. Any ambiguities and rating conflicts will be discussed relative to opportunities to refine the criteria to more clearly distinguish between ratings. If major changes to the risk-of-bias criteria are made based on the pilot phase (i.e., those that would likely result in revision of response), they will be documented in a protocol amendment along with the date modifications were made and the logic for the changes. It is also expected that information about confounding, exposure characterization, outcome assessment, and other important issues may be identified during or after data extraction, which can lead to further refinement of the risk-of-bias criteria (Sterne *et al.* 2014).

After assessors have independently made risk-of-bias determinations for a study across all risk-of-bias questions, the two assessors will compare their results to identify discrepancies and attempt to resolve them. Any remaining discrepancies will be considered by the project lead and, if needed, other members of the evaluation design team and/or technical advisors. The final risk-of-bias rating for each

Table 5: OHAT Risk-of-Bias Questions and Applicability by Study Design

Risk-of-Bias Questions	Experimental Animal*	Human Controlled Trials**	Cohort	Case-Control	Cross-Sectional***	Case Series
1. Was administered dose or exposure level adequately randomized?	X	X				
2. Was allocation to study groups adequately concealed?	X	X				
3. Did selection of study participants result in the appropriate comparison groups?			X	X	X	
4. Did study design or analysis account for important confounding and modifying variables?			X	X	X	X
5. Were experimental conditions identical across study groups?	X					
6. Were research personnel blinded to the study group during the study?	X	X				
7. Were outcome data complete without attrition or exclusion from analysis?	X	X	X	X	X	
8. Can we be confident in the exposure characterization?	X	X	X	X	X	X
9. Can we be confident in the outcome assessment (including blinding of outcome assessors)?	X	X	X	X	X	X
10. Were all measured outcomes reported?	X	X	X	X	X	X
11. Were there no other potential threats to internal validity?	X	X	X	X	X	X

*Experimental animal studies are controlled exposure studies. Non-human animal observational studies can be evaluated using the design features of observational human studies such as cross-sectional study design. **Human Controlled Trials are studies in humans with controlled exposure (e.g., Randomized Controlled Trials, non-randomized experimental studies)

***Cross-sectional studies include population surveys with individual data (e.g., NHANES) and surveys with aggregate data (i.e., ecological studies).

question will be recorded along with a statement of the basis for that rating or rationale in HAWC. The risk-of-bias assessment of included studies will be part of the study summaries released in materials for the draft state-of-the-science report that will be posted for public comment prior to peer review. Peer review will provide an opportunity for investigators and the public to comment on the risk-of-bias analysis.

Missing Information for Risk of Bias Assessment

OHAT will attempt to contact authors of included studies by email to obtain missing information considered critical for evaluating risk of bias that cannot be inferred from the study. If additional information or data are received from study authors, risk-of-bias judgments will be modified to reflect the updated study information. If OHAT does not receive a response from the authors by one month of the contact attempt, a risk of bias response of "NR" for "not reported; probably high risk of bias" will be used and a note made in the data extraction files that an attempt to contact the authors was unsuccessful.

STATE-OF-THE-SCIENCE REPORT FORMAT

The NTP state-of-the-science report on transgenerational inheritance of health effects associated with exposure to a wide range of stressors (e.g., environmental chemicals, drugs of abuse, nutrition and diet, pharmaceuticals, infectious agents, or stress) in humans or animals will include the following information:

Introduction

This section will provide a brief background on the topic.

Methodology

This section will provide a brief overview of the methodologies used in the review process, including:

- the research question
- the search strategy used to identify and retrieve studies
- the process for selecting the included studies
- the methods of data extraction
- the methods used to assess risk of bias of a sub-set of included studies (i.e., examples of human and animal studies)

Results

This section will include the results from the state-of-the-science evaluation on transgenerational inheritance of health effects associated with exposure to a wide range of stressors (e.g., environmental chemicals, drugs of abuse, nutrition and diet, pharmaceuticals, infectious agents, or stress) in humans or animals. Results will be presented in tables or figures as appropriate using HAWC. The results from the included studies will be discussed by exposure. This will include a description of:

- the number of studies identified that reported the various exposures and within the exposures the reported health outcomes reported
- full list of excluded studies, with reasons for exclusion documented for studies excluded at the full text review stage

- a summary of the results and risk-of-bias assessment to illustrate how risk-of-bias-assessment could be applied to studies of transgenerational study design. Risk-of-bias assessment will be evaluated from the 2 most extensive bodies of evidence for both human and animal studies. In other words, exposures for which there are the largest number of publications addressing the same health outcome so that risk-of-bias assessment can be compared across multiple studies.

Discussion

The discussion will provide a summary of the review findings, the areas of consistency and areas of uncertainty in bodies of evidence, including a discussion of any gaps identified in the evidence and any suggestions of areas for further research. Any important limitations of the review will be described and their impact on the available evidence will be discussed.

REFERENCES

Aiken CE, Ozanne SE. 2014. Transgenerational developmental programming. *Hum Reprod Update* 20(1): 63-75.

Barker DJ, Eriksson JG, Forsen T, Osmond C. 2002. Fetal origins of adult disease: strength of effects and biological basis. *Int J Epidemiol* 31(6): 1235-1239.

CLARITY Group at McMaster University. 2013. *Tools to assess risk of bias in cohort studies, case control studies, randomized controlled trials, and longitudinal symptom research studies aimed at the general population*. Available: http://www.evidencepartners.com/resources/ [accessed 15 January 2013].

Grossniklaus U, Kelly WG, Ferguson-Smith AC, Pembrey M, Lindquist S. 2013. Transgenerational epigenetic inheritance: how important is it? *Nat Rev Genet* 14(3): 228-235.

Guerrero-Bosagna C, Skinner MK. 2012. Environmentally induced epigenetic transgenerational inheritance of phenotype and disease. *Mol Cell Endocrinol* 354(1-2): 3-8.

Higgins J, Green S. 2011. Cochrane Handbook for Systematic Reviews of Interventions, The Cochrane Collaboration.

Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JA, Cochrane Bias Methods G, Cochrane Statistical Methods G. 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 343: d5928.

Moher D, Liberati A, Tetzlaff J, Altman DG. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *J Clin Epidemiol* 62(10): 1006-1012.

NIEHS. 2012. *Transgenerational Inheritance in Mammals After Environmental Exposure (TIME) (R01)(RFA-ES-12-006)*. Available: http://grants.nih.gov/grants/guide/rfa-files/RFA-ES-12-006.html.

Sterne J, Higgins J, Reeves B, on behalf of the development group for ACROBAT-NRSI. 2014. *A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI), Version 1.0.0*. Available: http://www.riskofbias.info [accessed 28 September 2014].

ABOUT THE PROTOCOL

Contributors

Evaluation Team

Evaluation teams are composed of federal staff and contractor staff. Contractor staff members are screened for potential conflicts of interest. Federal staff members are asked to do a self-evaluation for potential conflict of interest using the same criteria. Epidemiologists and toxicologists on OHAT evaluation teams have at least three years' experience and/or training in reviewing studies, including summarizing studies and critical review (e.g., assessing study quality and interpreting findings). Experience in evaluating occupational or environmental studies is preferred. Team members generally have at least a master's degree or equivalent level of experience in epidemiology, toxicology, environmental health sciences, or a related field.

Name	Affiliation
Vickie Walker	NIEHS/NTP, Project Lead
Abee Boyles, PhD	NIEHS/NTP
Andrew Rooney, PhD	NIEHS/NTP
Chad Blystone, PhD	NIEHS/NTP
Fred Tyson, PhD	NIEHS/DERT
Jerry Heindel, PhD	NIEHS/DERT
Katie Pelch, PhD	NIEHS/NTP
Kembra Howdeshell, PhD	NIEHS/NTP
Kristina Thayer, PhD	NIEHS/NTP
Kyla Taylor, MS	NIEHS/NTP
Lisa Chadwick, PhD	NIEHS/DERT
Michael Devito, PhD	NIEHS/NTP
Paul Wade, PhD	NIEHS/DIR
Retha Newbold	NIEHS/NTP, contractor
Stephanie Holmgren	NIEHS/NIH
Vicki Sutherland, PhD	NIEHS/NTP

Contract support: Will assist in data extraction and risk-of-bias assessment

Name	Affiliation
Pam Hartman	ICF Incorporated
Isaac Warren	ICF Incorporated

Technical Advisors

Technical advisors are outside experts retained on an as-needed basis to provide individual advice to the NTP for a specific topic. If technical advisors are selected for this project, they will be selected for their experience with transgenerational research or study design, specific chemicals or exposures relevant to the evaluation, or health effects of interest. Technical advisors would be screened for conflict of interest

prior to their service. Service as a technical advisor does not necessarily indicate that the advisor has read the entire protocol or endorses the final state-of-the-science document.

Peer Reviewers

The peer reviewers are outside experts selected for their experience with transgenerational research or study design and systematic review procedures. Peer reviewers will be screened for conflict of interest prior to their service. Service as a peer reviewer does not necessarily indicate that the reviewer endorses the final document or protocol.

Review of protocol:

Name	Affiliation
Juleen Lam, PhD	University of California at San Francisco
Johanna Rochester, PhD	TEDX, The Endocrine Disruption Exchange

Sources of Support

National Institute of Environmental Health Sciences/Division of the National Toxicology Program

Protocol History and Revisions

Date	Activity or revision
May 28, 2015:	**Draft evaluation protocol reviewed:** sent to peer reviewers for comment/review
June 24, 2015	**Evaluation protocol posted on NTP website**

APPENDICES

Appendix 1. PubMed Literature Search Strategy

Literature Search Method for PubMed	
#1	Transgeneration*[tiab] OR trans-generation*[tiab]
#2	multigeneration*[tiab] OR multi-generation*[tiab] OR intergeneration*[tiab] OR inter-generation*[tiab]
#3	(F2 OR "F 2" OR "F(2)" OR F3 OR "F 3" OR "F(3)" OR F4 OR "F 4" OR "F(4)" OR F5 OR "F 5" OR "F(5)" OR F6 OR "F 6" OR "F(6)" OR F7 OR "F 7" OR "F(7)" OR F8 OR "F 8" OR "F(8)" OR F9 OR "F 9" OR "F(9)" OR F10 OR "F 10" OR "F(10)") AND (generation*[tiab] OR offspring[tiab] OR progeny[tiab])
#4	"two generations"[tiab] OR "second generation"[tiab] OR "second generations"[tiab] OR "three generations"[tiab] OR "third generation"[tiab] OR "third generations"[tiab] OR "four generations"[tiab] OR "fourth generation"[tiab] OR "fourth generations"[tiab] OR "five generations"[tiab] OR "fifth generation"[tiab] OR "fifth generations"[tiab]
#5	Greatgrandparent*[tiab] OR "Great Grandparent"[tiab] OR "Great Grandparents"[tiab] OR "Great Grand Parent"[tiab] OR "Great Grand Parents"[tiab] OR Greatgrandfather*[tiab] OR "Great Grandfather"[tiab] OR "Great Grandfathers"[tiab] OR "Great Grand Father"[tiab] OR "Great Grand Fathers"[tiab] OR Greatgrandmother*[tiab] OR "Great Grandmother"[tiab] OR "Great Grandmothers"[tiab] OR "Great Grand Mother"[tiab] OR "Great Grand Mothers"[tiab] OR Greatgrandchild*[tiab] OR "Great Grandchild"[tiab] OR "Great Grandchildren"[tiab] OR "Great Grand Child"[tiab] OR "Great Grand Children"[tiab] OR Greatgrandson*[tiab] OR "Great Grandson"[tiab] OR "Great Grandsons"[tiab] OR "Great Grand Son"[tiab] OR "Great Grand Sons"[tiab] OR Greatgranddaughter*[tiab] OR "Great Granddaughter"[tiab] OR "Great Granddaughters"[tiab] OR "Great Grand Daughter"[tiab] OR "Great Grand Daughters"[tiab] OR Grandparent* [tiab] OR "Grand Parent"[tiab] OR "Grand Parents"[tiab] OR Grandfather* [tiab] OR "Grand Father"[tiab] OR "Grand Fathers"[tiab] OR Grandmother*[tiab] OR "Grand Mother"[tiab] OR "Grand Mothers"[tiab] OR Grandchild*[tiab] OR "Grand Child"[tiab] OR "Grand Children"[tiab] OR Granddaughter*[tiab] OR "Grand Daughter"[tiab] OR "Grand Daughters"[tiab] OR Grandson*[tiab] OR "Grand Son"[tiab] OR "Grand Sons"[tiab]
Full search	Full search was the inclusive combination of each of the concepts listed above (#1 AND #2 AND #3 AND #4 AND #5)

Appendix 2. Data Extraction Elements for Human Studies

HUMAN	
Funding	Funding source(s)
	Reporting of conflict of interest (COI) by authors (*reporting bias)
Subjects	Study population name/description
	Dates of study and sampling time frame
	Geography (country, region, state, etc.)
	Demographics (sex, race/ethnicity, age or lifestage at exposure and at outcome assessment)
	Number of subjects (target, enrolled, n per group in analysis, and participation/follow-up rates) (*missing data bias)
	Inclusion/exclusion criteria/recruitment strategy (*selection bias)
	Description of reference group (*selection bias)
Methods	Study design (e.g., prospective or retrospective cohort, nested case-control study, cross-sectional, population-based case-control study, intervention, case report, etc.)
	Length of follow-up (*information bias)
	Health outcome category, e.g., cardiovascular
	Health outcome, e.g., blood pressure (*reporting bias)
	Diagnostic or methods used to measure health outcome (*information bias)
	Confounders or modifying factors and how considered in analysis (e.g., included in final model, considered for inclusion but determined not needed (*confounding bias)
	Substance name and CAS number
	Exposure assessment (e.g., blood, urine, hair, air, drinking water, job classification, residence, administered treatment in controlled study, etc.) (*information bias)
	Methodological details for exposure assessment (e.g., HPLC-MS/MS, limit of detection) (*information bias)
	Statistical methods (*information bias)
Results	Exposure levels (e.g., mean, median, measures of variance as presented in paper, such as SD, SEM, 75th/90th/95th percentile, minimum/maximum); range of exposure levels, number of exposed cases
	Statistical findings (e.g., adjusted β, standardized mean difference, adjusted odds ratio, standardized mortality ratio, relative risk, etc.) or description of qualitative results. When possible, OHAT will convert measures of effect to a common metric with associated 95% confidence intervals (CI). Most often, measures of effect for continuous data are expressed as mean difference, standardized mean difference, and percent control response. Categorical data are typically expressed as odds ratio, relative risk (RR, also called risk ratio), or β values, depending on what metric is most commonly reported in the included studies and on OHAT's ability to obtain information for effect conversions from the study or through author query.

	If not presented in the study, statistical power can be assessed during data extraction using an approach that can detect a 10% to 20% change from response by control or referent group for continuous data, or a relative risk or odds ratio of 1.5 to 2 for categorical data, using the prevalence of exposure or prevalence of outcome in the control or referent group to determine sample size. For categorical data where the sample sizes of exposed and control or referent groups differ, the sample size of the exposed group will be used to determine the relative power category. Recommended sample sizes to achieve 80% power for a given effect size, i.e., 10% or 20% change from control, will be compared to sample sizes used in the study to categorize statistical power as "appears to be adequately powered" (sample size for 80% power met), somewhat underpowered (sample size is 75% to < 100% of number required for 80% power), "underpowered" (sample size is 50% to < 75% of number required for 80% power), or "severely underpowered" (sample size is < 50% of number required for 80% power).
	Observations on dose response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, non-monotonic)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.
Items marked with an asterisk (*) are examples of items that can be used to assess internal validity/risk of bias	

Appendix 3. Data Extraction Elements for Animal Studies

ANIMAL	
Funding	Funding source(s)
	Reporting of COI by authors (*reporting bias)
Animal Model	Sex
	Species
	Strain
	Source of animals
	Age or lifestage at start of dosing and at health outcome assessment
	Diet and husbandry information (e.g., diet name/source)
Treatment	Chemical name and CAS number
	Source of chemical
	Purity of chemical (*information bias)
	Dose levels or concentration (as presented and converted to mg/kg bw/d when possible)
	Other dose-related details, such as whether administered dose level was verified by measurement, information on internal dosimetry (*information bias)
	Vehicle used for exposed animals
	Route of administration (e.g., oral, inhalation, dermal, injection)
	Duration and frequency of dosing (e.g., hours, days, weeks when administration was ended, days per week)
Methods	Study design (e.g., single treatment, acute, subchronic (e.g., 90 days in a rodent), chronic, multigenerational, developmental, other)
	Guideline compliance (i.e., use of EPA, OECD, NTP or another guideline for study design, conducted under GLP guideline conditions, non-GLP but consistent with guideline study, non-guideline peer-reviewed publication)
	Number of animals per group (and dams per group in developmental studies) (*missing data bias)
	Randomization procedure, allocation concealment, blinding during outcome assessment (*selection bias)
	Method to control for litter effects in developmental studies (*information bias)
	Use of negative controls and whether controls were untreated, vehicle-treated, or both
	Report on data from positive controls – was expected response observed? (*information bias)
	Endpoint health category (e.g., reproductive)
	Endpoint (e.g., infertility)
	Diagnostic or method to measure endpoint (*information bias)
	Statistical methods (*information bias)
Results	Measures of effect at each dose or concentration level (e.g., mean, median, frequency, and measures of precision or variance) or description of qualitative results. When possible, OHAT will convert measures of effect to a common metric with associated 95% confidence intervals (CI). Most often, measures of effect for continuous data will be expressed as mean difference, standardized mean difference, and percent control response. Categorical data will be expressed as relative risk (RR, also called risk ratio).

	No Observed Effect Level (NOEL), Lowest Observed Effect Level (LOEL), benchmark dose (BMD) analysis, statistical significance of other dose levels, or other estimates of effect presented in paper. **Note:** The NOEL and LOEL are highly influenced by study design, do not give any quantitative information about the relationship between dose and response, and can be subject to author's interpretation (e.g., a statistically significant effect may not be considered biologically important). Also, a NOEL does not necessarily mean zero response. Ideally, the response rate at specific dose levels is used as the primary measure to characterize the response.
	If not presented in the study, statistical power can be assessed during data extraction using an approach that assesses the ability to detect a 10% to 20% change from control group's response for continuous data, or a relative risk or odds ratio of 1.5 to 2 for categorical data, using the outcome frequency in the control group to determine sample size. Recommended sample sizes to achieve 80% power for a given effect size, i.e., 10% or 20% change from control, will be compared to sample sizes used in the study to categorize statistical power as "appears to be adequately powered" (sample size for 80% power met), "somewhat underpowered" (sample size is 75% to < 100% of number required for 80% power), "underpowered" (sample size is 50% to < 75% of number required for 80% power), or "severely underpowered" (sample size is < 50% of number required for 80% power).
	Observations on dose response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, non-monotonic)
	Data on internal concentration, toxicokinetics, or toxicodynamics (when reported)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.
Items marked with an asterisk (*) are examples of items that can be used to assess internal validity/risk of bias	

Appendix 4. Risk-of-Bias Criteria

The OHAT risk-of-bias tool for human and animal studies (version date January 2015 and available at http://ntp.niehs.nih.gov/go/38673) reflects OHAT's current best practices and provides the detailed discussion and instructions for the risk-of-bias practices used in this evaluation. The OHAT tool uses a single set of questions (also called "elements" or "domains") to assess risk of bias across various study types to facilitate consideration of conceptually similar potential sources of bias across the human and animal evidence streams with a common terminology. Individual risk-of-bias questions are designated as only applicable to certain study designs (e.g., cohort studies or experimental animal studies), and a subset of the questions apply to each study design (**Table 5**).

The specific criteria used to assess risk of bias for this evaluation are outlined below for Human/observational studies and experimental animal studies. Based on preliminary literature searches we do not expect any controlled exposure studies in humans (i.e., human controlled trials) and therefore have not included risk-of-bias criteria for that study design. If relevant human controlled trials are identified, the criteria from the January 2015 OHAT risk-of-bias tool will be used to evaluate risk of bias.

Observational Studies (Human studies or wildlife animal studies)

Cohort studies

1. Was administered dose or exposure level adequately randomized? [NA]

2. Was allocation to study groups adequately concealed? [NA]

3. Did selection of study participants result in the appropriate comparison groups?

Definitely Low Risk of Bias (++)
• Direct evidence that subjects (both exposed and non-exposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, • **Note:** A study will be considered low risk of bias if baseline characteristics of groups differed but these differences were considered as potential confounding or stratification variables (see question #4),
Probably Low Risk of Bias (+)
• Indirect evidence that subjects (both exposed and non-exposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, • **OR** differences between groups would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that subjects (both exposed and non-exposed) were not similar, recruited within very different time frames, or had the very different participation/response rates, • **OR** there is insufficient information provided about the comparison group including a different rate of non-response without an explanation (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that subjects (both exposed and non-exposed) were not similar, recruited within very different time frames, or had the very different participation/response rates.

4. Did study design or analysis account for important confounding and modifying variables?

Definitely Low Risk of Bias (++)
• Direct evidence that appropriate adjustments or explicit considerations were made for the variables listed below as potential confounders and/or effect measure modifiers in the final analyses through the use of statistical models to reduce research-specific bias including standardization, matching, adjustment in multivariate model, stratification, propensity scoring, or other methods that were appropriately justified. Acceptable consideration of appropriate adjustment factors includes cases when the factor is not included in the final adjustment model because the author conducted analyses that indicated it did not need to be included, • **AND** there is direct evidence that primary covariates and confounders were assessed using valid and reliable measurements, • **AND** there is direct evidence that other exposures anticipated to bias results were not present or were appropriately measured and adjusted for. In occupational studies or studies of contaminated sites, other chemical exposures known to be associated with those settings were appropriately considered. • **Note:** The following variables should be considered as potential confounders and/or effect measure modifiers for the relationship between exposure and outcomes: age, sex, race/ethnicity, smoking, body mass index, and variables that represent socioeconomic status (e.g., educational level, household income). During the pilot phase for the risk-of-bias assessment, additional variables will be considered for the two specific exposure scenarios (holocaust and radiation) in this evaluation based on prior reports of the variables associations with both exposure levels and outcomes measured in the studies. If major changes to the risk-of-bias criteria are made based on the pilot phase (i.e., those that would likely result in revision of response), they will be documented in a protocol amendment along with the date modifications were made and the logic for the changes.
Probably Low Risk of Bias (+)
• Indirect evidence that appropriate adjustments were made, • **OR** it is deemed that not considering or only considering a partial list of covariates or confounders in the final analyses would not appreciably bias results, • **AND** there is evidence (direct or indirect) that covariates and confounders considered were assessed using valid and reliable measurements, • **OR** it is deemed that the measures used would not appreciably bias results (i.e., the authors justified the validity of the measures from previously published research), • **AND** there is evidence (direct or indirect) that other co-exposures anticipated to bias results were not present or were appropriately adjusted for, • **OR** it is deemed that co-exposures present would not appreciably bias results. • **Note:** this includes insufficient information provided on co-exposures in general population studies.

Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the distribution of important covariates and known confounders differed between the groups and was not appropriately adjusted for in the final analyses, • **OR** there is insufficient information provided about the distribution of known confounders (record "NR" as basis for answer), • **OR** there is indirect evidence that covariates and confounders considered were assessed using measurements of unknown validity, • **OR** there is insufficient information provided about the measurement techniques used to assess covariates and confounders considered (record "NR" as basis for answer), • **OR** there is indirect evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for, • **OR** there is insufficient information provided about co-exposures in occupational studies or studies of contaminated sites where high exposures to other chemical exposures would have been reasonably anticipated (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the distribution of important covariates and known confounders differed between the groups, confounding was demonstrated, and was not appropriately adjusted for in the final analyses, • **OR** there is direct evidence that covariates and confounders considered were assessed using non valid measurements, • **OR** there is direct evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for.

5. Were experimental conditions identical across study groups? [NA]

6. Were the research personnel blinded to the study group during the study? [NA]

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)
• Direct evidence that loss of subjects (i.e., incomplete outcome data) was adequately addressed and reasons were documented when human subjects were removed from a study. • **Note**: Acceptable handling of subject attrition includes: very little missing outcome data; reasons for missing subjects unlikely to be related to outcome (for survival data, censoring unlikely to be introducing bias); missing outcome data balanced in numbers across study groups, with similar reasons for missing data across groups, • **OR** missing data have been imputed using appropriate methods and characteristics of subjects lost to follow up or with unavailable records are described in identical way and are not significantly different from those of the study participants.
Probably Low Risk of Bias (+)
• Indirect evidence that loss of subjects (i.e., incomplete outcome data) was adequately addressed and reasons were documented when human subjects were removed from a study, • **OR** it is deemed that the proportion lost to follow-up would not appreciably bias results. This would include reports of no statistical differences in characteristics of subjects lost to follow up or with unavailable records from those of the study participants. Generally, the higher the ratio of participants with missing data to participants with events, the greater potential there is for bias. For studies with a long duration of follow-up, some withdrawals are inevitable.

Probably High Risk of Bias (-) or (NR)
• Indirect evidence that loss of subjects (i.e., incomplete outcome data) was unacceptably large and not adequately addressed, • **OR** there is insufficient information provided about numbers of subjects lost to follow-up (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that loss of subjects (i.e., incomplete outcome data) was unacceptably large and not adequately addressed. • **Note**: Unacceptable handling of subject attrition includes: reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across study groups; or potentially inappropriate application of imputation.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)
• Direct evidence that exposure was consistently assessed (i.e., under the same method and time-frame) using well-established methods that directly measure exposure (e.g., measurement of radiation exposure to individuals), • **OR** exposure was assessed using less-established methods that directly measure exposure and are validated against well-established methods, • **AND** exposure was assessed in a relevant time-window for development of the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes, • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished • **Note:** During the pilot phase for the risk-of-bias assessment, additional consideration will be given to the methods for exposure characterization of the two specific exposure scenarios (holocaust and radiation) in this evaluation are being explored. If major changes to the risk-of-bias criteria are made based on the pilot phase (i.e., those that would likely result in revision of response), they will be documented in a protocol amendment along with the date modifications were made and the logic for the changes.
Probably Low Risk of Bias (+)
• Indirect evidence that the exposure was consistently assessed using well-established methods that directly measure exposure), • **OR** exposure was assessed using indirect measures (e.g., questionnaire or occupational exposure assessment by a certified industrial hygienist) that have been validated or empirically shown to be consistent with methods that directly measure exposure (i.e., inter-methods validation: one method vs. another), • **AND** exposure was assessed in a relevant time-window for development of the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes (at a minimum from high exposure or ever exposed from low exposure or never exposed), • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.

Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the exposure was assessed using poorly validated methods that directly measure exposure, • **OR** there is evidence that the exposure was assessed using indirect measures that have not been validated or empirically shown to be consistent with methods that directly measure exposure (e.g., questionnaire, job-exposure matrix or self-report without validation) (record "NR" as basis for answer), • **OR** there is insufficient information provided about the exposure assessment, including validity and reliability, but no evidence for concern about the method used (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the exposure was assessed using methods with poor validity, • **OR** evidence of exposure misclassification (e.g., differential recall of self-reported exposure).

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)
• Direct evidence that the outcome was assessed using well-established methods (e.g., gold standard), • **AND** subjects had been followed for the same length of time in all study groups, • **AND** there is direct evidence that the outcome assessors were adequately blinded to the study group or exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes. • **NOTE:** Well-established methods will depend on the outcome, but examples of such methods may include: diagnostic methods using commercial kits, commercial laboratories, or standard assays such as ELISAs with sufficiently low variation and limits of detection to allow discrimination between groups (or evidence that the assay could have detected a difference based on responses to a positive control)
Probably Low Risk of Bias (+)
• Indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable but not the gold standard), • **AND** subjects had been followed for the same length of time in all study groups • **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, • **AND** there is indirect evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes, • **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results, which is more likely to apply to objective outcome measures, • **NOTE:** Acceptable, but not ideal assessment methods will depend on the outcome, but examples of such methods may include proxy reporting of outcomes and mining of data collected for other purposes.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the outcome assessment method is an insensitive instrument (e.g., a questionnaire used to assess outcomes with no information on validation), • **OR** the length of follow up differed by study group, • **OR** there is indirect evidence that it was possible for outcome assessors (including study subjects if outcomes were self-reported) to infer the study group prior to reporting outcomes, • **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).

Definitely High Risk of Bias (--)
• Direct evidence that the outcome assessment method is an insensitive instrument, • **OR** the length of follow up differed by study group, • **OR** there is direct evidence for lack of adequate blinding of outcome assessors (including study subjects if outcomes were self-reported), including no blinding or incomplete blinding.

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance. Note: this is a standard to ensure clear reporting and this project will not include a meta-analysis.
Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, • **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

This question is used to address project-specific issues. We are not aware of risk-of-bias considerations for observational studies with a transgenerational study design, but would add the specific question here and mark it as a modification to the protocol if such an issue is identified during the project. Note for experimental studies, consideration of the litter as the experimental unit is evaluated in this question, but there is no equivalent consideration for observational studies. This question will also be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.

Cross Sectional and Case Series Studies

1. Was administered dose or exposure level adequately randomized? [NA]

2. Was allocation to study groups adequately concealed? [NA]

3. Did selection of study participants result in the appropriate comparison groups? [NA for Case series]

Definitely Low Risk of Bias (++)
• Direct evidence that subjects (both exposed and non-exposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, • **Note:** A study will be considered low risk of bias if baseline characteristics of groups differed but these differences were considered as potential confounding or stratification variables (see question #4),
Probably Low Risk of Bias (+)
• Indirect evidence that subjects (both exposed and non-exposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, • **OR** differences between groups would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that subjects (both exposed and non-exposed) were not similar, recruited within very different time frames, or had the very different participation/response rates, • **OR** there is insufficient information provided about the comparison group including a different rate of non-response without an explanation (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that subjects (both exposed and non-exposed) were not similar, recruited within very different time frames, or had the very different participation/response rates.

4. Did study design or analysis account for important confounding and modifying variables?

Definitely Low Risk of Bias (++)
• Direct evidence that appropriate adjustments or explicit considerations were made for the variables listed below as potential confounders and/or effect measure modifiers in the final analyses through the use of statistical models to reduce research-specific bias including standardization, matching, adjustment in multivariate model, stratification, propensity scoring, or other methods that were appropriately justified. Acceptable consideration of appropriate adjustment factors includes cases when the factor is not included in the final adjustment model because the author conducted analyses that indicated it did not need to be included, • **AND** there is direct evidence that primary covariates and confounders were assessed using valid and reliable measurements, • **AND** there is direct evidence that other exposures anticipated to bias results were not present or were appropriately measured and adjusted for. In occupational studies or studies of contaminated sites, other chemical exposures known to be associated with those settings were appropriately considered. • **Note:** The following variables should be considered as potential confounders and/or effect measure modifiers for the relationship between exposure and outcomes: age, sex, race/ethnicity, smoking, body mass index, and variables that represent socioeconomic status (e.g., educational level, household income). During the pilot phase for the risk of bias assessment, additional variables will be considered for the two specific exposure scenarios (holocaust and radiation) in this evaluation based on prior reports of the variables associations with both exposure levels and outcomes measured in the studies. If major changes to the risk-of-bias criteria are made based on the pilot phase (i.e., those that would likely result in revision of response), they will be documented in a protocol amendment along with the date modifications were made and the logic for the changes.
Probably Low Risk of Bias (+)
• Indirect evidence that appropriate adjustments were made, • **OR** it is deemed that not considering or only considering a partial list of covariates or confounders in the final analyses would not appreciably bias results, • **AND** there is evidence (direct or indirect) that covariates and confounders considered were assessed using valid and reliable measurements, • **OR** it is deemed that the measures used would not appreciably bias results (i.e., the authors justified the validity of the measures from previously published research), • **AND** there is evidence (direct or indirect) that other co-exposures anticipated to bias results were not present or were appropriately adjusted for, • **OR** it is deemed that co-exposures present would not appreciably bias results. • **Note:** this includes insufficient information provided on co-exposures in general population studies.

Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the distribution of important covariates and known confounders differed between the groups and was not appropriately adjusted for in the final analyses, • **OR** there is insufficient information provided about the distribution of known confounders (record "NR" as basis for answer), • **OR** there is indirect evidence that covariates and confounders considered were assessed using measurements of unknown validity, • **OR** there is insufficient information provided about the measurement techniques used to assess covariates and confounders considered (record "NR" as basis for answer), • **OR** there is indirect evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for, • **OR** there is insufficient information provided about co-exposures in occupational studies or studies of contaminated sites where high exposures to other chemical exposures would have been reasonably anticipated (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the distribution of important covariates and known confounders differed between the groups, confounding was demonstrated, and was not appropriately adjusted for in the final analyses, • **OR** there is direct evidence that covariates and confounders considered were assessed using non valid measurements, • **OR** there is direct evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for.

5. Were experimental conditions identical across study groups? [NA]

6. Were the research personnel blinded to the study group during the study? [NA]

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)
• Direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably Low Risk of Bias (+)
• Indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that exclusion of subjects from analyses was not adequately addressed, • **OR** there is insufficient information provided about why subjects were removed from the study or excluded from analyses (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that exclusion of subjects from analyses was not adequately addressed. • **Note**: Unacceptable handling of subject exclusion from analyses includes: reason for exclusion likely to be related to true outcome, with either imbalance in numbers or reasons for exclusion across study groups.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)
• Direct evidence that exposure was consistently assessed (i.e., under the same method and time-frame) using well-established methods that directly measure exposure (e.g., measurement of radiation exposure to individuals), • **OR** exposure was assessed using less-established methods that directly measure exposure and are validated against well-established methods, • **AND** exposure was assessed in a relevant time-window for development of the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes, • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished. • **Note:** During the pilot phase for the risk-of-bias assessment, additional consideration will be given to the methods for exposure characterization of the two specific exposure scenarios (holocaust and radiation) in this evaluation. If major changes to the risk-of-bias criteria are made based on the pilot phase (i.e., those that would likely result in revision of response), they will be documented in a protocol amendment along with the date modifications were made and the logic for the changes.
Probably Low Risk of Bias (+)
• Indirect evidence that the exposure was consistently assessed using well-established methods that directly measure exposure), • **OR** exposure was assessed using indirect measures (e.g., questionnaire or occupational exposure assessment by a certified industrial hygienist) that have been validated or empirically shown to be consistent with methods that directly measure exposure (i.e., inter-methods validation: one method vs. another), • **AND** exposure was assessed in a relevant time-window for development of the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes (at a minimum from high exposure or ever exposed from low exposure or never exposed), • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the exposure was assessed using poorly validated methods that directly measure exposure • **OR** there is evidence that the exposure was assessed using indirect measures that have not been validated or empirically shown to be consistent with methods that directly measure exposure (e.g., a job-exposure matrix or self-report without validation) (record “NR” as basis for answer), • **OR** there is insufficient information provided about the exposure assessment, including validity and reliability, but no evidence for concern about the method used (record “NR” as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the exposure was assessed using methods with poor validity, • **OR** evidence of exposure misclassification (e.g., differential recall of self-reported exposure).

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)
• Direct evidence that the outcome was assessed using well-established methods (e.g., gold standard), • **AND** subjects had been followed for the same length of time in all study groups, • **AND** there is direct evidence that the outcome assessors were adequately blinded to the study group or exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes. • **NOTE:** Well-established methods will depend on the outcome, but examples of such methods may include: diagnostic methods using commercial kits, commercial laboratories, or standard assays such as ELISAs with sufficiently low variation and limits of detection to allow discrimination between groups (or evidence that the assay could have detected a difference based on responses to a positive control)
Probably Low Risk of Bias (+)
• Indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable but not the gold standard), • **AND** subjects had been followed for the same length of time in all study groups • **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, • **AND** there is indirect evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes, • **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results, which is more likely to apply to objective outcome measures, • **NOTE:** Acceptable, but not ideal assessment methods will depend on the outcome, but examples of such methods may include proxy reporting of outcomes and mining of data collected for other purposes.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the outcome assessment method is an insensitive instrument, • **OR** there is indirect evidence that it was possible for outcome assessors to infer the exposure level prior to reporting outcomes (including that subjects self-reporting outcomes were likely aware of reported links between the exposure and outcome), • **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the outcome assessment method is an insensitive instrument, • **OR** there is direct evidence that outcome assessors were aware of the exposure level prior to reporting outcomes (including if subjects self-reporting outcomes were aware of reported links between the exposure and outcome).

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance. Note: this is a standard to ensure clear reporting and this project will not include a meta-analysis.

Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, • **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

This question is used to address project-specific issues. We are not aware of risk-of-bias considerations for observational studies with a transgenerational study design, but would add the specific question here and mark it as a modification to the protocol if such an issue is identified during the project. Note for experimental studies, consideration of the litter as the experimental unit is evaluated in this question, but there is no equivalent consideration for observational studies. This question will also be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.

Case Control Studies

1. Was administered dose or exposure level adequately randomized? [NA]

2. Was allocation to study groups adequately concealed? [NA]

3. Did selection of study participants result in the appropriate comparison groups?

Definitely Low Risk of Bias (++)
• Direct evidence that cases and controls were similar (e.g., recruited from the same eligible population including being of similar age, gender, ethnicity, and eligibility criteria other than outcome of interest as appropriate), recruited within the same time frame, and controls are described as having no history of the outcome, • **Note:** A study will be considered low risk of bias if baseline characteristics of groups differed but these differences were considered as potential confounding or stratification variables (see question #4)
Probably Low Risk of Bias (+)
• Indirect evidence that cases and controls were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age), recruited within the same time frame, and controls are described as having no history of the outcome, • **OR** it is deemed differences between cases and controls would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that controls were drawn from a very dissimilar population than cases or recruited within very different time frames, • **OR** there is insufficient information provided about the appropriateness of controls including rate of response reported for cases only (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that controls were drawn from a very dissimilar population than cases or recruited within very different time frames.

4. Did study design or analysis account for important confounding and modifying variables?

Definitely Low Risk of Bias (++)
• Direct evidence that appropriate adjustments were made for the variables listed below as potential confounders and/or effect measure modifiers in the final analyses through the use of statistical models to reduce research-specific bias including standardization, matching of cases and controls, adjustment in multivariate model, stratification, propensity scoring, or other methods were appropriately justified, • **AND** there is direct evidence that primary covariates and confounders were assessed using valid and reliable measurements, • **AND** there is direct evidence that other exposures anticipated to bias results were not present or were appropriately measured and adjusted for. • **Note:** The following variables should be considered as potential confounders and/or effect measure modifiers for the relationship between exposure and outcomes: age, sex, race/ethnicity, smoking, body mass index, and variables that represent socioeconomic status (e.g., educational level, household income). During the pilot phase for the risk-of-bias assessment, additional variables will be considered for the two specific exposure scenarios (holocaust and radiation) in this evaluation based on prior reports of the variables associations with both exposure levels and outcomes measured in the studies. If major changes to the risk-of-bias criteria are made based on the pilot phase (i.e., those that would likely result in revision of response), they will be documented in a protocol amendment along with the date modifications were made and the logic for the changes.

Probably Low Risk of Bias (+)
• Indirect evidence that appropriate adjustments were made, • **OR** it is deemed that not considering or only considering a partial list of covariates or confounders in the final analyses would not appreciably bias results, • **AND** there is evidence (direct or indirect) that covariates and confounders considered were assessed using valid and reliable measurements, • **OR** it is deemed that the measures used would not appreciably bias results (i.e., the authors justified the validity of the measures from previously published research), • **AND** there is evidence (direct or indirect) that other co-exposures anticipated to bias results were not present or were appropriately adjusted for, • **OR** it is deemed that co-exposures present would not appreciably bias results. • **Note:** this includes insufficient information provided on co-exposures in general population studies.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the distribution of important covariates and known confounders differed between cases and controls and was not investigated further, • **OR** there is insufficient information provided about the distribution of known confounders in cases and controls (record "NR" as basis for answer), • **OR** there is indirect evidence that covariates and confounders considered were assessed using measurements of unknown validity, • **OR** there is insufficient information provided about the measurement techniques used to assess covariates and confounders considered (record "NR" as basis for answer), • **OR** there is indirect evidence that there was an unbalanced provision of additional co-exposures across cases and controls, which were not appropriately adjusted for, • **OR** there is insufficient information provided about co-exposures in occupational studies or studies of contaminated sites where high exposures to other chemical exposures would have been reasonably anticipated (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the distribution of important covariates and known confounders differed between cases and controls, confounding was demonstrated, but was not appropriately adjusted for in the final analyses, • **OR** there is direct evidence that covariates and confounders considered were assessed using non valid measurements, • **OR** there is direct evidence that there was an unbalanced provision of additional co-exposures across cases and controls, which were not appropriately adjusted for.

5. Were experimental conditions identical across study groups? [NA]

6. Were the research personnel blinded to the study group during the study? [NA]

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)
• Direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably Low Risk of Bias (+)
• Indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

Probably High Risk of Bias (-) or (NR)
• Indirect evidence that exclusion of subjects from analyses was not adequately addressed, • **OR** there is insufficient information provided about why subjects were removed from the study or excluded from analyses (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that exclusion of subjects from analyses was not adequately addressed. • **Note**: Unacceptable handling of subject exclusion from analyses includes: reason for exclusion likely to be related to true outcome, with either imbalance in numbers or reasons for exclusion across study groups.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)
• Direct evidence that exposure was consistently assessed (i.e., under the same method and time-frame) using well-established methods that directly measure exposure (e.g., measurement of radiation exposure to individuals), • **OR** exposure was assessed using less-established methods that directly measure exposure and are validated against well-established methods, • **AND** exposure was assessed in a relevant time-window for development of the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes, • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished. • **Note:** During the pilot phase for the risk-of-bias assessment, additional consideration will be given to the methods for exposure characterization of the two specific exposure scenarios (holocaust and radiation) in this evaluation. If major changes to the risk-of-bias criteria are made based on the pilot phase (i.e., those that would likely result in revision of response), they will be documented in a protocol amendment along with the date modifications were made and the logic for the changes.
Probably Low Risk of Bias (+)
• Indirect evidence that the exposure was consistently assessed using well-established methods that directly measure exposure), • **OR** exposure was assessed using indirect measures (e.g., questionnaire or occupational exposure assessment by a certified industrial hygienist) that have been validated or empirically shown to be consistent with methods that directly measure exposure (i.e., inter-methods validation: one method vs. another), • **AND** exposure was assessed in a relevant time-window for development of the outcome, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes (at a minimum from high exposure or ever exposed from low exposure or never exposed), • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.

Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the exposure was assessed using poorly validated methods that directly measure exposure, • **OR** there is direct evidence that the exposure was assessed using indirect measures that have not been validated or empirically shown to be consistent with methods that directly measure exposure (e.g., a job-exposure matrix or self-report without validation) (record "NR" as basis for answer), • **OR** there is insufficient information provided about the exposure assessment, including validity and reliability, but no evidence for concern about the method used (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the exposure was assessed using methods with poor validity, • **OR** evidence of exposure misclassification (e.g., differential recall of self-reported exposure).

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)
• Direct evidence that the outcome was assessed using well-established methods (e.g., gold standard), • **AND** subjects had been followed for the same length of time in all study groups, • **AND** there is direct evidence that the outcome assessors were adequately blinded to the study group or exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes. • **NOTE:** Well-established methods will depend on the outcome, but examples of such methods may include: diagnostic methods using commercial kits, commercial laboratories, or standard assays such as ELISAs with sufficiently low variation and limits of detection to allow discrimination between groups (or evidence that the assay could have detected a difference based on responses to a positive control)
Probably Low Risk of Bias (+)
• Indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable but not the gold standard), • **AND** subjects had been followed for the same length of time in all study groups • **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, • **AND** there is indirect evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes, • **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results, which is more likely to apply to objective outcome measures, • **NOTE:** Acceptable, but not ideal assessment methods will depend on the outcome, but examples of such methods may include proxy reporting of outcomes and mining of data collected for other purposes.

Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the outcome was assessed in cases (i.e., case definition) using an insensitive instrument, • **OR** there is insufficient information provided about how cases were identified (record "NR" as basis for answer). • **OR** there is indirect evidence that it was possible for outcome assessors to infer the exposure level prior to reporting outcomes (including that subjects self-reporting outcomes were likely aware of reported links between the exposure and outcome), • **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the outcome was assessed in cases (i.e., case definition) using an insensitive instrument, • **OR** there is direct evidence that outcome assessors were aware of the exposure level prior to reporting outcomes (including that subjects self-reporting outcomes were aware of reported links between the exposure and outcome).

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance.
Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, • **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not). Note: this is a standard to ensure clear reporting and this project will not include a meta-analysis.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as basis for answer).

Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

This question is used to address project-specific issues. We are not aware of risk-of-bias considerations for observational studies with a transgenerational study design, but would add the specific question here and mark it as a modification to the protocol if such an issue is identified during the project. Note for experimental studies, consideration of the litter as the experimental unit is evaluated in this question, but there is no equivalent consideration for observational studies. This question will also be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.

Experimental Animal Studies

1. Was administered dose or exposure level adequately randomized?

Definitely Low Risk of Bias (++)
• Direct evidence that animals were allocated to any study group including controls using a method with a random component, • **AND** there is direct evidence that the study used a concurrent control group as an indication that randomization covered all study groups, • **Note:** Acceptable methods of randomization include: referring to a random number table, using a computer random number generator, coin tossing, or shuffling cards (Higgins and Green 2011). • **Note:** Restricted randomization (e.g., blocked randomization) to ensure particular allocation ratios will be considered low bias. Similarly, stratified randomization approaches that attempt to minimize imbalance between groups on important prognostic factors (e.g., body weight) will be considered acceptable.
Probably Low Risk of Bias (+)
• Indirect evidence that animals were allocated to any study group including controls using a method with a random component (i.e., authors state random allocation, without description of method), • **AND** evidence that the study used a concurrent control group as an indication that randomization covered all study groups, • **OR** it is deemed that allocation without a clearly random component would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that animals were allocated to study groups using a method with a non-random component, • **OR** indirect evidence that there was a lack of a concurrent control group, • **OR** there is insufficient information provided about how animals were allocated to study groups (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that animals were allocated to study groups using a non-random method including judgment of the investigator, the results of a laboratory test or a series of tests, • **OR** direct evidence that there was a lack of a concurrent control group.

2. Was allocation to study groups adequately concealed?

Definitely Low Risk of Bias (++)
• Direct evidence that at the time of assigning study groups the research personnel did not know what group animals were allocated to, and it is unlikely that they could have broken the blinding of allocation until after assignment was complete and irrevocable. • **Note:** Acceptable methods used to ensure allocation concealment include sequentially numbered treatment containers of identical appearance or equivalent methods.
Probably Low Risk of Bias (+)
• Indirect evidence that at the time of assigning study groups the research personnel did not know what group animals were allocated to and it is unlikely that they could have broken the blinding of allocation until after assignment was complete and irrevocable, • **OR** it is deemed that lack of adequate allocation concealment would not appreciably bias results.

Probably High Risk of Bias (-) or (NR)
• Indirect evidence that at the time of assigning study groups it was possible for the research personnel to know what group animals were allocated to, or it is likely that they could have broken the blinding of allocation before assignment was complete and irrevocable, • **OR** there is *insufficient* information provided about allocation to study groups (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that at the time of assigning study groups it was possible for the research personnel to know what group animals were allocated to, or it is likely that they could have broken the blinding of allocation before assignment was complete and irrevocable.

3. Did selection of study participants result in the appropriate comparison groups? [NA]

4. Did study design or analysis account for important confounding and modifying variables? [NA]

5. Were experimental conditions identical across study groups?

Definitely Low Risk of Bias (++)
• Direct evidence that same vehicle was used in control and experimental animals, • **AND** direct evidence that non-treatment-related experimental conditions were identical across study groups (i.e., the study report explicitly provides this level of detail).
Probably Low Risk of Bias (+)
• Indirect evidence that the same vehicle was used in control and experimental animals, • **OR** it is deemed that the vehicle used would not appreciably bias results, • **AND** identical non-treatment-related experimental conditions are assumed if authors did not report differences in housing or husbandry.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the vehicle differed between control and experimental animals, • **OR** authors did not report the vehicle used (record "NR" as basis for answer), • **OR** there is indirect evidence that non-treatment-related experimental conditions were not comparable between study groups.
Definitely High Risk of Bias (--)
• Direct evidence from the study report that control animals were untreated, or treated with a different vehicle than experimental animals, • **OR** there is direct evidence that non-treatment-related experimental conditions were not comparable between study groups.

6. Were the research personnel blinded to the study group during the study?

Definitely Low Risk of Bias (++)
• Direct evidence that the research personnel were adequately blinded to study group, and it is unlikely that they could have broken the blinding during the study. Methods used to ensure blinding include central allocation; sequentially numbered treatment containers of identical appearance; sequentially numbered animal cages; or equivalent methods.

Probably Low Risk of Bias (+)
• Indirect evidence that the research personnel were adequately blinded to study group, and it is unlikely that they could have broken the blinding during the study, • **OR** it is deemed that lack of adequate blinding during the study would not appreciably bias results. This would include cases where blinding was not possible but research personnel took steps to minimize potential bias, such as restricting the knowledge of study group to veterinary or supervisory personnel monitoring for overt toxicity, or randomized husbandry or handling practices (e.g., placement in the animal room, necropsy order, etc.).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the research personnel were not adequately blinded to study group, • **OR** there is insufficient information provided about blinding to study group during the study (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the research personnel were not adequately blinded to study group.

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)
• Direct evidence that loss of animals was adequately addressed and reasons were documented when animals were removed from a study. • **Note**: Acceptable handling of attrition includes: very little missing outcome data; reasons for missing animals unlikely to be related to outcome (or for survival data, censoring unlikely to be introducing bias); missing outcome data balanced in numbers across study groups, with similar reasons for missing data across groups; missing outcomes is not enough to impact the effect estimate. • **OR** missing data have been imputed using appropriate methods (insuring that characteristics of animals are not significantly different from animals retained in the analysis).
Probably Low Risk of Bias (+)
• Indirect evidence that loss of animals was adequately addressed and reasons were documented when animals were removed from a study, • **OR** it is deemed that the proportion lost would not appreciably bias results. This would include reports of no statistical differences in characteristics of animals removed from the study from those remaining in the study.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that loss of animals was unacceptably large and not adequately addressed, • **OR** there is insufficient information provided about loss of animals (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that loss of animals was unacceptably large and not adequately addressed. • **Note**: Unacceptable handling of attrition or exclusion includes: reason for loss is likely to be related to true outcome, with either imbalance in numbers or reasons for loss across study groups.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)
• Direct evidence that the exposure to vinclozolin (including purity and stability) was independently characterized and purity confirmed generally as ≥98%, (and compliance with the treatment, if applicable), • **AND** that exposure was consistently administered (i.e., with the same method and time-frame) across treatment groups, • **AND** for dietary or drinking water studies that information is provided on consumption or internal dose metrics to confirm expected exposure levels sufficiently to allow discrimination between exposure groups, • **AND** if internal dose metrics are available, there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished, • **AND** if internal dose metrics are available, the study used spiked samples to confirm assay performance. • **Note:** if internal dose measurements are made, measurement of serum or whole-blood vinclozolin is the standard accepted biomarker of exposure (preferred over urine or feces) using quantitative techniques such as liquid chromatography-triple quadrupole mass spectrometry (LC-MS/MS) and high pressure liquid chromatography with tandem mass spectrometry (HPLC/MS). Additional considerations for exposure characterization of the two specific exposure scenarios (vinclozolin and radiation) in this evaluation are being explored.
Probably Low Risk of Bias (+)
• Direct evidence that the exposure to vinclozolin (including purity and stability) was independently characterized and purity confirmed generally as ≥98% (i.e., the supplier of the chemical provides documentation of the purity of the chemical), (and compliance with the treatment, if applicable) • **OR** direct evidence that purity was independently confirmed as ≥95% and it is deemed that impurities of up to 5% would not appreciably bias results, • **AND** that exposure was consistently administered (i.e., with the same method and time-frame) across treatment groups, • **AND** for dietary or drinking water studies no information is provided on consumption or internal dose metrics, • **AND** if internal dose metrics are available, there is indirect evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the exposure (including purity of the test substance and compliance with the treatment, if applicable) was assessed using poorly validated methods, • **OR** there is insufficient information provided about the validity of the exposure assessment method, but no evidence for concern (record “NR” as basis for answer), • **AND** if internal dose metrics are available, there is indirect evidence that most of the exposure data measurements are below the limit of quantitation for the assay such that different exposure groups cannot be distinguished.
Definitely High Risk of Bias (--)
• Direct evidence that the exposure (including purity of the test substance and compliance with the treatment, if applicable) was assessed using poorly validated methods.

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)
• Direct evidence that the outcome was assessed using well-established methods (e.g., gold standard) • **AND** assessed at the same length of time after initial exposure in all study groups, • **AND** there is direct evidence that the outcome assessors were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes. • **NOTE:** Well-established methods will depend on the outcome, but examples of such methods may include: objectively measured antibody or cytokine concentrations with diagnostic methods using commercial kits, commercial laboratories with experience in the assay, or standard assays such as ELISAs for IgG and with sufficiently low variation and limits of detection to allow discrimination of responses between treatment groups (or direct evidence that the assay could have detected a difference based on responses to a positive control). Additional considerations for outcome assessment may be added before the protocol is finalized (e.g., to focus on reproductive health outcomes which are one of the common endpoints).
• Indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable but not the gold standard), • **AND** assessed at the same length of time after initial exposure in all study groups, • **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, • **AND** there is indirect evidence that the outcome assessors were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes, • **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results, which is more likely to apply to objective outcome measures. • **NOTE** For some outcomes, particularly histopathology assessment, outcome assessors are not blind to study group as they require comparison to the control to appropriately judge the outcome, but additional measures such as multiple levels of independent review by trained pathologists can minimize potential bias. • **NOTE** Acceptable assessment methods will depend on the outcome, but examples of such methods may include: objectively measured antibody or cytokine concentrations with diagnostic methods using commercial kits with some variation, but ability to discriminate between the high dose treatment and control group (or indirect evidence that the assay could have detected a difference based on responses to a positive control).
• Indirect evidence that the outcome assessment method is an insensitive instrument, • **OR** the length of time after initial exposure differed by study group, • **OR** there is indirect evidence that it was possible for outcome assessors to infer the study group prior to reporting outcomes without sufficient quality control measures, • **OR** there is insufficient information provided about blinding of outcome assessors (record “NR” as basis for answer).
• Direct evidence that the outcome assessment method is an insensitive instrument, • **OR** the length of time after initial exposure differed by study group, • **OR** there is direct evidence for lack of adequate blinding of outcome assessors, including no blinding or incomplete blinding without quality control measures.

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance. Note: this is a standard to ensure clear reporting and this project will not include a meta-analysis.
Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, • **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as answer basis).
Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

This question is used to address project-specific issues. Because transgenerational studies involve exposure and then tracking health effects over multiple generations, consideration of the litter effect is important. It is recognized that fetuses from a given litter tend to exhibit a similar response to a chemical exposure. Therefore, the litter must be considered the experimental unit for study design and statistical analysis. This question will also be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.